

AC P09799; (Rel.10, created)
DT 01-MAR-1989 (Rel.10, last sequence update)
DE 15-JUL-1999 (Rel.38, last annotation update)
ID VICILIN GC72-A PRECURSOR (ALPHA-GLOBULIN A).
OS Gossypium hirsutum (Upland cotton).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
CC Magnoliophyta; eudicotyledons: core eudicots.; Rosidae; eurosids II;
CC Malvales; Malvaceae; Gossypium.
RN [1]
RP SEQUENCE FROM N.A.
RA Chlan C.A., Borrito K., Kamalay J.A., Dure L. III;
RT "Developmental biochemistry of cottonseed embryogenesis and
RT germination. XIX. Sequences and genomic organization of the alpha
RT globulin (vicilin) genes of cottonseed."
RL Plant Mol. Biol. 9:533-546(1987).
CC -1- FUNCTION: SEED STORAGE PROTEIN.
CC -1- SUBCELLULAR LOCATION: CYTOSOL;ER MEMBRANE-BOUND VACUOLAR PROTEIN
CC BODIES.
CC -1- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
CC CONVICILIN, CONGLYCININ, ETC.).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M19378; AAA33069.1; .
DR PIR: S06398; S06398.
DR HSSP: P50477; ICAX.
DR INTERPRO: IPR01113; .
DR PfAM: PF00546; Seedstore_7s; 1.
KW Seed storage protein; signal.
FT SIGNAL
FT CHAIN 1 23
SQ SEQUENCE 605 AA; 71049 MW; C9DB9371C976953B CRC64;

Query Match 76.7%; Score 188; DB 1; Length 605;
Best Local Similarity 75.0%; Pred. No.2,5e-13;
Matches 30; Conservative 7; Mismatches 3; Indels 0; Gaps 0;

DY 3 ROFOECQCHHQEOQRPEKKQCVRECKRYOENPMKGER 42
 :::||||| | |||||:|||||:|||||:|||||:
DB 120 KQFKECGQRCQGWQEQRPERKQQCVKCEGQYQEDPMKGER 159

RESULT 3
TRX_DROVI STANDARD: PRT; 3828 AA.

AC Q24742;
DT 01-OCT-2000 (Rel.40, Created)
DT 01-OCT-2000 (Rel.40, Last sequence update)
DT 01-OCT-2000 (Rel.40, Last annotation update)
DE TRITHORAX PROTEIN.
GN TRX.
OS Drosophila virilis (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
CC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
CC Ephydroidea; Drosophilidae; Drosophila.
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLIN=96100387; PubMed=8555104;
RA Tiliib S., Sedkov Y., Mizrokh L., Mazo A.;
RT "Conservation of structure and expression of the trithorax gene
RT between Drosophila virilis and Drosophila melanogaster".
RL Mech. Dev. 53:113-122(1995).
CC -1- FUNCTION: FUNCTIONS IN SEGMENT DETERMINATION THROUGH INTERACTION
CC WITH GENES OF BRITHORAX (BX-C) AND ANTENNAPEDIA (ANT-X) COMPLEXES.
CC IT CAN BEHAVE AS AN ACTIVATOR OF BX-C.

```

CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- SIMILARITY: BELONGS TO THE TITINORAX FAMILY OF TRANSCRIPTION
CC FACTORS.
CC -1- SIMILARITY: CONTAINS 1 'SET' DOMAIN.
CC -1- SIMILARITY: CONTAINS 5 PHD-FINGER DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC
CC EMBL; Z50038; CAA90349.1; -.
CC DR HSSP; P04002; IMF4.
CC DR FLVBASE; FB9N001484; DVLR\trx.
CC DR INTERPRO; IPR001214; -.
CC DR INTERPRO; IPR001965; -.
CC DR PFAM; PF00628; PHD; 2.
CC DR PFAM; PF00856; SET; 1.
CC KW Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;
CC Nuclear protein; Developmental protein; Activator.
CC FT DOMAIN 1251 1334 PHD 1.
CC FT DOMAIN 1335 1380 PHD 2.
CC FT DOMAIN 1408 1469 PHD 3.
CC FT DOMAIN 1708 1767 PHD 4 (ATYPICAL).
CC FT DOMAIN 1768 1818 PHD 5 (ATYPICAL).
CC FT DOMAIN 3701 3810 SET DOMAIN.
CC FT DOMAIN 28 41 POLY-ALA.
CC FT DOMAIN 66 71 POLY-ASP.
CC FT DOMAIN 160 164 POLY-ASP.
CC FT DOMAIN 173 182 POLY-ALA.
CC FT DOMAIN 221 228 POLY-GLN.
CC FT DOMAIN 243 251 POLY-ALA.
CC FT DOMAIN 253 258 POLY-THR.
CC FT DOMAIN 292 296 POLY-ALA.
CC FT DOMAIN 538 546 POLY-ASP.
CC FT DOMAIN 1072 1075 POLY-GLU.
CC FT DOMAIN 2483 3271 GLN-RICH.
CC FT DOMAIN 3333 3339 POLY-ASP.
CC SQ SEQUENCE 3828 AA; 413721 MW; 320596CF303A3C504 CRC64;

Query Match 30.2%; Score 74; DB 1; Length 3828;
Best Local Similarity 38.2%; Pred. No. 2.1;
Matches 13; Conservative 14; Mismatches 7; Indels 0; Gaps 0;

QY 2 QRFQECQOHCHQOEQRPEKQGCVCBEREKYQE 35
Db 2995 QOEPEQOHLHQOQOQOQOQOQHMQOHOQOQOQ 3028

RESULT 4
ID ALL1_SIN1AL STANDARD; PRT; 127 AA.
AC P15322;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE ALLERGEN SIN A 1, SMALL AND LARGE CHAINS (SIN A 1).
OS Sinapis alba (White mustard) (Brassicaceae).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Sinapis.
OC [1]
RP SEQUENCE.
RC TISSUE=SEED;
RX MEDLINE=99030681; Pubmed=3181153;
RA Mendez-Arias L., Moneo I., Dominguez J., Rodriguez R.,
FT "Primary structure of the major allergen of yellow mustard (Sinapis
FT alba L.) seed, Sin a I.";
FT Eur. J. Biochem. 177:159-166(1988)

```

```
CC -1- FUNCTION: THIS IS A 2S SEED STORAGE PROTEIN.
CC -1- SUBUNIT: THE PROTEIN CONSISTS OF TWO CHAINS LINKED BY DISULFIDE
CC BONDS.
CC -1- DISEASE: ALLERGIC DISEASE, CABBAGE ALLERGY.
CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
DR PIR: S01792; S01792.
DR PIR: S01791; S01791.
DR INTERPRO: IPR000617; -.
DR PRAM: PF01631; Seedstore_2S; 1.
DR PRINTS: PR00496; NAPIIN.
KM Allergen; Seed storage protein.
FT CHAIN 1 39 SMALL CHAIN.
FT NON_CONS 39 40
FT CHAIN 40 127 LARGE CHAIN.
FT VARIANT 6 6 R -> G.
FT SEQUENCE 127 AA; 14180 MW; 4CD920284F04EEFD CRC64;
SQ
```

```
Query Match 29.0%; Score 71; DB 1; Length 127;
Best Local Similarity 32.7%; Pred. No. 0.21;
Matches 18; Conservative 7; Mismatches 8; Indels 22; Gaps 3;
```

```
Db 2 0ROFOE-----COQHCHQO-----EQRPKKQOCVRCRCREKYQENP 37
11 RKEPQAOHLRACQOQWLHKQAMQSGSPQGPQGRPLLLQOC---CNELHQBEP 62

RESULT 5
ALL1_BRAJU STANDARD; PRT; 129 AA.
AC P80207; P80215;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE ALLERGEN BRA J 1-E, SMALL AND LARGE CHAINS (BRA J 1).
OS Brassica juncea (leaf mustard) (Indian mustard).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Brassica.
RN [1]
RP TISSUE=SEED.
RC MEDLINE=93356721; PubMed=7688955;
RX Monsalve R.I., Gonzalez de la Pena M.A., Menendez-Arias L.,
RA Lopez-Otin C., Villalba M., Rodriguez R.;
RT "Characterization of a new oriental-mustard (Brassica juncea)
allergen, Bra j 1E: detection of an allergenic epitope.";
RL Biochem. J. 293:625-632(1993).
CC -1- FUNCTION: THIS IS A 2S SEED STORAGE PROTEIN.
CC -1- SUBUNIT: THE MATURE PROTEIN CONSISTS OF A SMALL AND A LARGE CHAIN
CC LINKED BY TWO DISULFIDE BONDS.
CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
DR INTERPRO: IPR000617; -.
DR PRAM: PF01631; Seedstore_2S; 1.
DR PRINTS: PR00496; NAPIIN.
KM Allergen; Seed storage protein.
FT CHAIN 1 37 SMALL CHAIN.
FT NON_CONS 37 38
FT CHAIN 38 129 LARGE CHAIN.
FT VARIANT 6 6 R -> I.
FT VARIANT 20 20 F -> K.
FT SEQUENCE 129 AA; 14644 MW; D6F28E03F62B08F8 CRC64;
SQ
```

```
Query Match 29.0%; Score 71; DB 1; Length 129;
Best Local Similarity 32.7%; Pred. No. 0.21;
Matches 18; Conservative 7; Mismatches 8; Indels 22; Gaps 3;
```

```
Db 2 0ROFOE-----COQHCHQO-----EQRPKKQOCVRCRCREKYQENP 37
11 RKEPQAOHLRACQOQWLHKQAMQSGSPQGPQGRPLLLQOC---CNELHQBEP 61
```

```
RESULT 6
ITRY_SINAR STANDARD; PRT; 130 AA.
AC P38057;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE TRYPSIN INHIBITOR (TISA).
OS Sinapis arvensis (Charlock).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Sinapis.
RN [1]
RP TISSUE=SEED.
RC MEDLINE=94350545; PubMed=8070965;
RX Svendsen I.B., Nicolova D., Goshv I., Genov N.;
RT "Primary structure, spectroscopic and inhibitory properties of a two-
chain trypsin inhibitor from the seeds of charlock (Sinapis arvensis
L), a member of the napin protein family.";
RT Int. J. Pept. Protein Res. 43:425-430(1994).
RL Int. J. Pept. Protein Res. 43:425-430(1994).
CC -1- FUNCTION: INHIBITS TRYPSIN WITH A KI OF 7 X 10(-6) M.
CC -1- SUBUNIT: THE PROTEIN CONSISTS OF TWO CHAINS LINKED BY DISULFIDE
CC BONDS.
CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
DR INTERPRO: IPR000617; -.
DR PRAM: PF01631; Seedstore_2S; 1.
DR PRINTS: PR00496; NAPIIN.
KM Seed storage protein; Multigene family; Polymorphism.
FT CHAIN 1 39 SMALL CHAIN.
FT NON_CONS 39 40
FT CHAIN 40 130 LARGE CHAIN.
FT VARIANT 32 32 R -> M.
FT VARIANT 73 73 A -> S.
FT VARIANT 77 77 K -> R.
FT VARIANT 81 81 Q -> R.
FT VARIANT 87 87 H -> Q.
FT VARIANT 89 89 Q -> H.
FT VARIANT 91 91 G -> Q.
FT VARIANT 97 97 E -> Q.
FT VARIANT 98 98 I -> V.
FT VARIANT 99 99 R -> S.
FT VARIANT 106 106 T -> K.
FT VARIANT 123 123 N -> Q.
FT VARIANT 124 124 K -> G.
FT VARIANT 126 126 M -> V.
FT VARIANT 53 57 MISSING (IN FORM II).
FT SEQUENCE 130 AA; 14682 MW; EC02E4B26D18D0F2 CRC64;
SQ
```

```
Query Match 29.0%; Score 71; DB 1; Length 130;
Best Local Similarity 32.7%; Pred. No. 0.21;
Matches 18; Conservative 7; Mismatches 8; Indels 22; Gaps 3;
```

```
Db 2 0ROFOE-----COQHCHQO-----EQRPKKQOCVRCRCREKYQENP 37
11 RKEPQAOHLRACQOQWLHKQAMQSGSPQGPQGRPLLLQOC---CNELHQBEP 62
```

```
RESULT 7
LP61_EIMTE STANDARD; PRT; 255 AA.
ID LP61_EIMTE
AC P15714;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE ANTIGEN LPMC-61 (FRAGMENT).
OS Eimeria tenella.
OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;
OC Elmeria.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SPOROZOITE;
```

RA MEDLINE-90348718; PubMed-2200963;
 RA Ko C., Smith C.K. II, McDowell M.;
 RT "Identification and characterization of a target antigen of a
 RT monoclonal antibody directed against *Elmeria tenella* merozoites.";
 RL Mol. Biochem. Parasitol. 41:53-64(1990).
 CC -1- FUNCTION: UNKNOWN. THE GLN-RICH TANDEM REPEATS MAY BE IMPORTANT
 CC FOR AN UNKNOWN ASPECT OF THE PARASITIC LIFE CYCLE. MAY BE AN
 CC IMPORTANT IMMUNOGEN.
 CC -1- SUBUNIT: MAY BE COVALENTLY LINKED BY DISULFIDE BONDS TO OTHER
 CC POLYPEPTIDES TO FORM THE 80 KDA ANTIGEN.
 CC -1- DEVELOPMENTAL STAGE: PRESENT IN ALL STAGES THROUGHOUT THE
 CC SPOROULATION OF THE OOCYSTS AND IN THE SPOROZOITES FOLLOWING
 CC EXCISTATION.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M30933; AAA29079.1; -
 DR PIR: A60637; A60637.
 KM Antigen; Sporozoite; Repeat; Sporulation.
 FT NON_TER 1
 FT DOMAIN 1 210 12 X APPROXIMATE TANDEM REPEATS, GLN-
 FT RICH.
 FT REPEAT 18 48 1.
 FT REPEAT 49 57 2.
 FT REPEAT 58 65 3.
 FT REPEAT 66 78 4.
 FT REPEAT 79 90 5.
 FT REPEAT 91 103 6.
 FT REPEAT 104 140 7.
 FT REPEAT 141 152 8.
 FT REPEAT 153 164 9.
 FT REPEAT 165 172 10.
 FT REPEAT 173 192 11.
 FT REPEAT 193 210 12.
 FT NON_TER 255
 FT SEQUENCE 255 AA; 31267 MW; 8C5E6005FFFC2DB3 CRC64;
 SQ
 Query Match 26.9%; Score 66; DB 1; Length 255;
 Best Local Similarity 36.4%; Pred. No. 1.3;
 Matches 16; Conservative 8; Mismatches 14; Indels 6; Gaps 1;
 QY 1 SQRFQECQGHCHQ-----EQREPKKQCCVRECKEKQENPM 38
 DB 122 SQQQQLQQCGQQQQQQQLQQQQWSEQQQQQQQQQQWPEQQQQQQQQQ 165
 RESULT 8
 ATX1_HUMAN
 ID ATX1_HUMAN STANDARD; PRT; 816 AA.
 AC P54253;
 DT 01-OCT-1996 (Rel. 34; Created)
 DT 01-OCT-1996 (Rel. 34; Last sequence update)
 DT 15-JUL-1999 (Rel. 38; Last annotation update)
 DE ATAXIN-1 (SPINOCEREBELLAR ATAXIA TYPE 1 PROTEIN).
 GN SCAL.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=CEREBELLUM, AND BRAIN;
 RA MEDLINE-95038838; PubMed-7951322;
 RA Banfi S., Servadio A., Chung M.-Y., Kwiatkowski T.J. Jr., McCall A.E.,
 RA Duvick L.A., Shen Y., Roth E.J., Orr H.T., Zoghbi H.Y.;
 RT "Identification and characterization of the gene causing type 1
 RT spinocerebellar ataxia.";

RL Nat. Genet. 7:513-519(1994).
 CC -1- ALTERNATIVE PRODUCTS: AT LEAST TWO FORMS ARE PRODUCED BY
 CC ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED THROUGHOUT THE BODY.
 CC -1- POLYMORPHISM: THE POLY-GLN REGION OF SCAL IS HIGHLY POLYMORPHIC
 CC (9 TO 39 REPEATS) IN THE NORMAL POPULATION AND IS EXPANDED TO
 CC ABOUT 40-61 REPEATS IN SCAL PATIENTS. LONGER EXPANSIONS RESULT IN
 CC EARLIER ONSET AND MORE SEVERE CLINICAL MANIFESTATIONS OF THE
 CC DISEASE.
 CC -1- DISEASE: DEFECTS IN SCAL ARE THE CAUSE OF SPINOCEREBELLAR ATAXIA
 CC TYPE 1; ALSO KNOWN AS OLIVOPONTOCEREBELLAR ATROPHY I (OPA 1).
 CC SCAL IS AN AUTOSOMAL DOMINANT NEURODEGENERATIVE DISORDER
 CC CHARACTERIZED BY PROGRESSIVE NEURONAL LOSS IN THE CEREBELLUM,
 CC BRAIN STEM AND SPINOCEREBELLAR TRACTS. CLINICAL FEATURES ARE
 CC CEREBELLAR ATAXIA, DYSPARTHRIA, OPHTALMOPARESIS, MUSCLE WASTING AND
 CC NEUROPATHY. ONSET OF THE DISEASE USUALLY OCCURS IN THE THIRD OR
 CC FOURTH DECADE OF LIFE AND DEATH OCCURS TEN TO TWENTY YEARS LATER.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X79204; CAA55793.1; -
 DR MIM: 164400; -
 KM Polymorphism; Triplet repeat expansion; Alternative splicing.
 FT DOMAIN 197 226 POLY-GLN.
 FT SEQUENCE 816 AA; 87051 MW; D49BA5DB423D0777 CRC64;
 SQ
 Query Match 26.9%; Score 66; DB 1; Length 816;
 Best Local Similarity 52.2%; Pred. No. 3.8;
 Matches 12; Conservative 7; Mismatches 4; Indels 0; Gaps 0;
 QY 2 QRFQECQGHCHQEQREPKKQ 24
 DB 200 QQQQQQQQQQQHQQQQQQQQQQQQQQQQ 222
 RESULT 9
 YAK1_YEAST
 ID YAK1_YEAST STANDARD; PRT; 807 AA.
 AC P14680;
 DT 01-APR-1990 (Rel. 14; Created)
 DT 01-APR-1990 (Rel. 14; Last sequence update)
 DT 01-OCT-2000 (Rel. 40; Last annotation update)
 DE PROTEIN KINASE YAK1 (EC 2.7.1.1).
 GN YAK1 OR YJL141C OR J0652.
 OS Saccharomyces cerevisiae (Baker's Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
 OC Saccharomycetaceae; Saccharomyces.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE-90108683; PubMed-2558053;
 RA Garrett S., Broach J.;
 RT "Loss of Ras activity in *Saccharomyces cerevisiae* is suppressed by
 RT disruptions of a new kinase gene, YAK1, whose product may act
 RT downstream of the cAMP-dependent protein kinase.";
 RL Genes Dev. 3:1336-1348(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / FY1679;
 RA MEDLINE-96408771; PubMed-8813765;
 RA Katsoulou C., Tzeremla M., Tavernarakis N., Alexandraki D.;
 RT "Sequence analysis of a 40.7 kb segment from the left arm of yeast
 RT chromosome X reveals 14 known genes and 13 new open reading frames
 RT including homologues of genes clustered on the right arm of
 RT chromosome XI.";
 RL Yeast 12:787-797(1996).
 CC -1- FUNCTION: THIS PROTEIN OCCURS IN REVERTANT RAS/CAMP MUTANTS


```

CC DIFFERENT SPECIES.
CC -1- P1M: KNOWN SUBSTRATE OF TRANSGLUTAMINASE. SOME 200 ARGININES ARE
CC PROBABLY CONVERTED TO CITRULLINES BY PEPTIDYLARGININE DEIMINASE.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE S-100
CC FAMILY.
CC -1- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L09190; AAA65582.1; -.
CC PIR: A45973; A45973.
CC HSSP: P02633; 1B0C.
CC MTM: 190370; -.
CC INTERPRO: IPR001751; -.
CC INTERPRO: IPR002048; -.
CC PFAM: PF01023; S.100; 1.
CC PFAM: PF00036; ehand; 1.
CC PROSITE: PS00018; EF_HAND; 1.
CC PROSITE: PS00303; S100_CABP; 1.
CC Repeat; Calcium-binding.
CC FT DOMAIN 1 91 S-100 LIKE.
CC FT CA_BIND 22 33 SITE I (LOW AFFINITY) (POTENTIAL).
CC FT CA_BIND 62 73 SITE II (HIGH AFFINITY) (POTENTIAL).
CC FT DOMAIN 314 390 6 X 13 AA TANDM REPEATS OF
CC R-R-E-Q-E-E-R-R-E-Q-O-L.
CC FT REPEAT 314 326 1-1 (APPROXIMATE).
CC FT REPEAT 327 339 1-2 (APPROXIMATE).
CC FT REPEAT 340 351 1-3 (APPROXIMATE).
CC FT REPEAT 352 364 1-4.
CC FT REPEAT 355 377 1-5.
CC FT REPEAT 378 390 1-6.
CC FT DOMAIN 391 444 9 X 6 AA TANDM REPEATS OF R-R-E-Q-O-L.
CC FT REPEAT 391 396 2-1.
CC FT REPEAT 397 402 2-2.
CC FT REPEAT 403 408 2-3.
CC FT REPEAT 409 414 2-4.
CC FT REPEAT 415 420 2-5.
CC FT REPEAT 421 426 2-6.
CC FT REPEAT 427 432 2-7.
CC FT REPEAT 433 438 2-8.
CC FT REPEAT 439 444 2-9.
CC FT DOMAIN 444 702 9 X 28 AA APPROXIMATE TANDM REPEATS.
CC FT DOMAIN 923 1162 8 X 30 AA TANDM REPEATS.
CC FT REPEAT 923 952 4-1.
CC FT REPEAT 953 982 4-2.
CC FT REPEAT 983 1012 4-3.
CC FT REPEAT 1013 1042 4-4.
CC FT REPEAT 1043 1072 4-5.
CC FT REPEAT 1073 1102 4-6.
CC FT REPEAT 1103 1132 4-7.
CC FT REPEAT 1133 1162 4-8.
CC FT DOMAIN 1250 1849 23 X 26 AA APPROXIMATE TANDM REPEATS.
CC FT CONFLICT 1752 1752 F -> L (IN REF. 2).
CC FT CONFLICT 1857 1857 Q -> K (IN REF. 2).
CC FT CONFLICT 1880 1880 V -> G (IN REF. 2).
CC SEQID 1898 AA; 247219 MW; A74B5947EB62E31D CRC64;

```

Query Match 26.1%; Score 64; DB 1; Length 1898;
 Best Local Similarity 31.8%; Pred. No. 13;
 Matches 14; Conservative 14; Mismatches 12; Indels 4; Gaps 1;

2 QROFOECQCHQOEOPEKKQOCVRCREKRYO---ENPWRC 41
 1117 KRRQREKRCREBELOQEBOLNREKRRRQLEKRYOFR 1160

```

RESULT 14
2SS1_BRANA STANDARD; PRT; 110 AA.
ID 2SS1_BRANA
AC P24565;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DE NAPIN IA AND IB SMALL CHAIN AND LARGE CHAINS.
OS Brassica napus (Rape).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Brassica.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87308224; PubMed=3624251;
RA Josefsson L.-G., Lemman M., Ericson M.L., Rask L.;
RT "Structure of a gene encoding the 1.7 S storage protein, napin, from
RL Brassica napus.";
J. Biol. Chem. 262:12196-12201(1987).

```

Query Match 25.9%; Score 63.5; DB 1; Length 110;
 Best Local Similarity 33.3%; Pred. No. 1.2;
 Matches 17; Conservative 7; Mismatches 8; Indels 19; Gaps 3;

2 QROFOECQCHQOEOPEKKQOCVRCREKRYOEN 36
 6 QREFOEQHRLACQWIRQLAGSPGSGPQPMRLREC---CNLYQED 53

```

RESULT 15
2SS2_BRANA STANDARD; PRT; 178 AA.
ID 2SS2_BRANA
AC P01090;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-AUG-1992 (Rel. 23, Last annotation update)
DE NAPIN 2 PRECURSOR (1.7S SEED STORAGE PROTEIN).
OS Brassica napus (Rape).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Brassica.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87308224; PubMed=3624251;
RA Josefsson L.-G., Lemman M., Ericson M.L., Rask L.;
RT "Structure of a gene encoding the 1.7 S storage protein, napin, from
RL Brassica napus.";
J. Biol. Chem. 262:12196-12201(1987).

```

RN [2]
RP REVISIONS.
RA Josefsson L.-G.;
RL Submitted (JUL-1987) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=87033665; PubMed=3771543;
RA Ericson M.L., Roedin J., Lemman M., Glimelius K., Josefsson L.-G.,
RA Rask L.;
RT "Structure of the rapeseed 1.7 S storage protein, napin, and its
RT precursor";
RL J. Biol. Chem. 261:14576-14581(1986).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, TOWER;
RX MEDLINE=84113267; PubMed=6689334;
RA Crouch M.L., Tenberge K.M., Simon A.E., Perl R.;
RT "cDNA clones for Brassica napus seed storage proteins: evidence from
RT nucleotide sequence analysis that both subunits of napin are cleaved
RT from a precursor polypeptide.";
RL J. Mol. Appl. Genet. 2:273-283(1983).
CC -1- FUNCTION: THE SMALL, BASIC, WATER-SOLUBLE NAPINS ARE ONE OF THE
CC TWO MAJOR KINDS OF STORAGE PROTEINS SYNTHESIZED IN THE SEED DURING
CC ITS MATURATION.
CC -1- SUBUNIT: THE MATURE PROTEIN CONSISTS OF A SMALL AND A LARGE CHAIN
CC LINKED BY DISULFIDE BONDS.
CC -1- TISSUE SPECIFICITY: COTYLEDONS AND THE AXIS.
CC -1- SIMILARITY: BELONGS TO THE 2S SEED STORAGE ALBUMINS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: K01545; AAA33006.1; -;
DR EMBL: J02586; AAA32997.1; -;
DR EMBL: J02798; AAA87348.1; -;
DR PIR: A01329; NMRP2.
DR PIR: A29801; A29801.
DR PIR: A25997; A25997.
DR INTERPRO: IPR000617; -;
DR PIR: PF01631; SeedStore_25; 1.
DR PIR: PR00496; NAPIN.
KW Seed storage protein; Signal; Multigene family.
FT SIGNAL 1 21
FT PROPEP 22 38
FT CHAIN 39 74 SMALL CHAIN.
FT PROPEP 75 94
FT CHAIN 95 175 LARGE CHAIN.
FT CONFLICT 37 37 D -> N (IN REF. 4).
FT CONFLICT 76 76 S -> N (IN REF. 4).
SQ SEQUENCE 178 AA; 20104 MW; 734E561971B539FF CRC64;

Query Match 25.9%; Score 63.5; DB 1; Length 178;
Best Local Similarity 25.7%; Pred. No. 1.8;
Matches 18; Conservative 7; Mismatches 8; Indels 37; Gaps 3;

QY 2 QROFGE-----COQHCHQO-----EORPEKKOQCVR 27
:::11: 111 1:1 :111 11
DB 48 RKEFDQAOHLRACQOQWLRQAMQSGGSPWTLDFEFDFEDMENPQGPQORPPLLOQC-- 105

QY 28 ECREKYOENP 37
| | :11 |
DB 106 -CNEIHQEER 114

Search completed: March 1, 2001, 16:16:58
Job time: 438 sec